Cornell Bowers CIS
Information Science

Georgia Tech

Using LLMs to Analyze Incident Reporting and Support Policymaking: Opportunities and Challenges

José A. Guridi - Cornell University
Eunji Emily Kim - Georgia Institute of Technology

Generated with ChatGPT

AI Incidents and Governance

- Risks associated with AI deployment—ranging from bias and discrimination to unintended system failures—are becoming more apparent (Acemoglu, 2021; Weidinger et al., 2022).
- Countries are taking different approaches to address AI potential harms through policies and regulations (Elgesem, 2023; Lupo, 2023; OECD.AI, 2021).
- One area where regulations and policies are being developed is incident reporting (Lupo, 2023), which is seen as a potential tool to prevent repeated AI failures (McGregor, 2021).

Prior work on AI incidents

AI Incidents to understand ethical issues

Prior work has analyzed incidents database to understand the ethical issues covered (McGregor, 2021; Nasim et al., 2022; Stanley & Dorton, 2023; Wei & Zhou, 2022). Feffer et al. (2023) explores the role of AI incidents in raising awareness of AI harms.

AI Incidents databases and practice

Turri and Dzombak (2023) examine reporting techniques of incidents across different industries to improve AI incidents reporting.

AI Incidents for policy/regulation

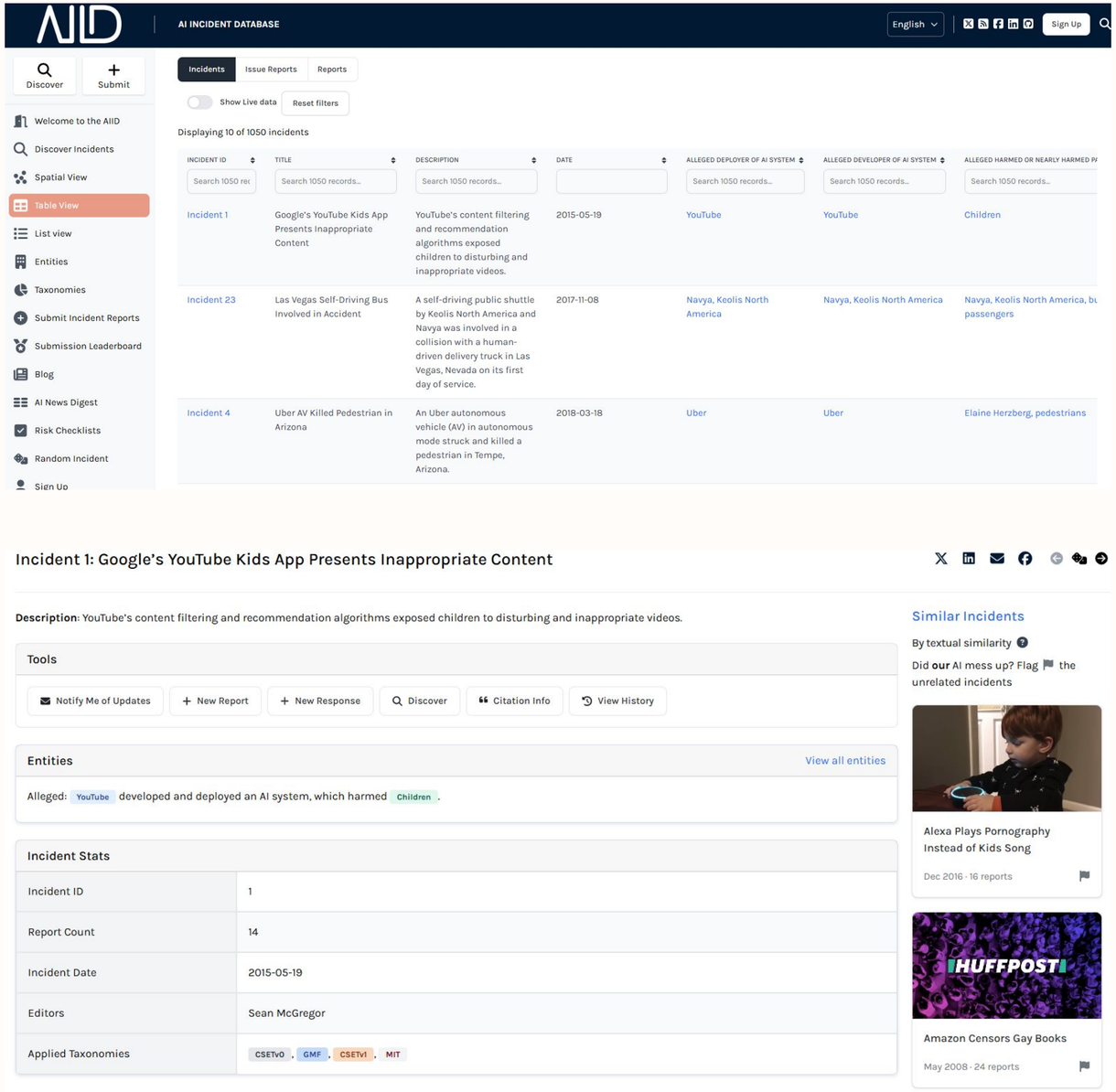
Lupo (2023) explores AI incidents and their relation with different areas of AI regulation.

Taxonomy of Risks

Slattery et al. (2025) created an AI risk repository by systematically reviewing AI incidents and AI risk taxonomies and creating a high-level Causal Taxonomy of AI Risks and a mid-level domain taxonomy of AI risk. This taxonomy is incorporated in the AI Incidents Database (AIID).

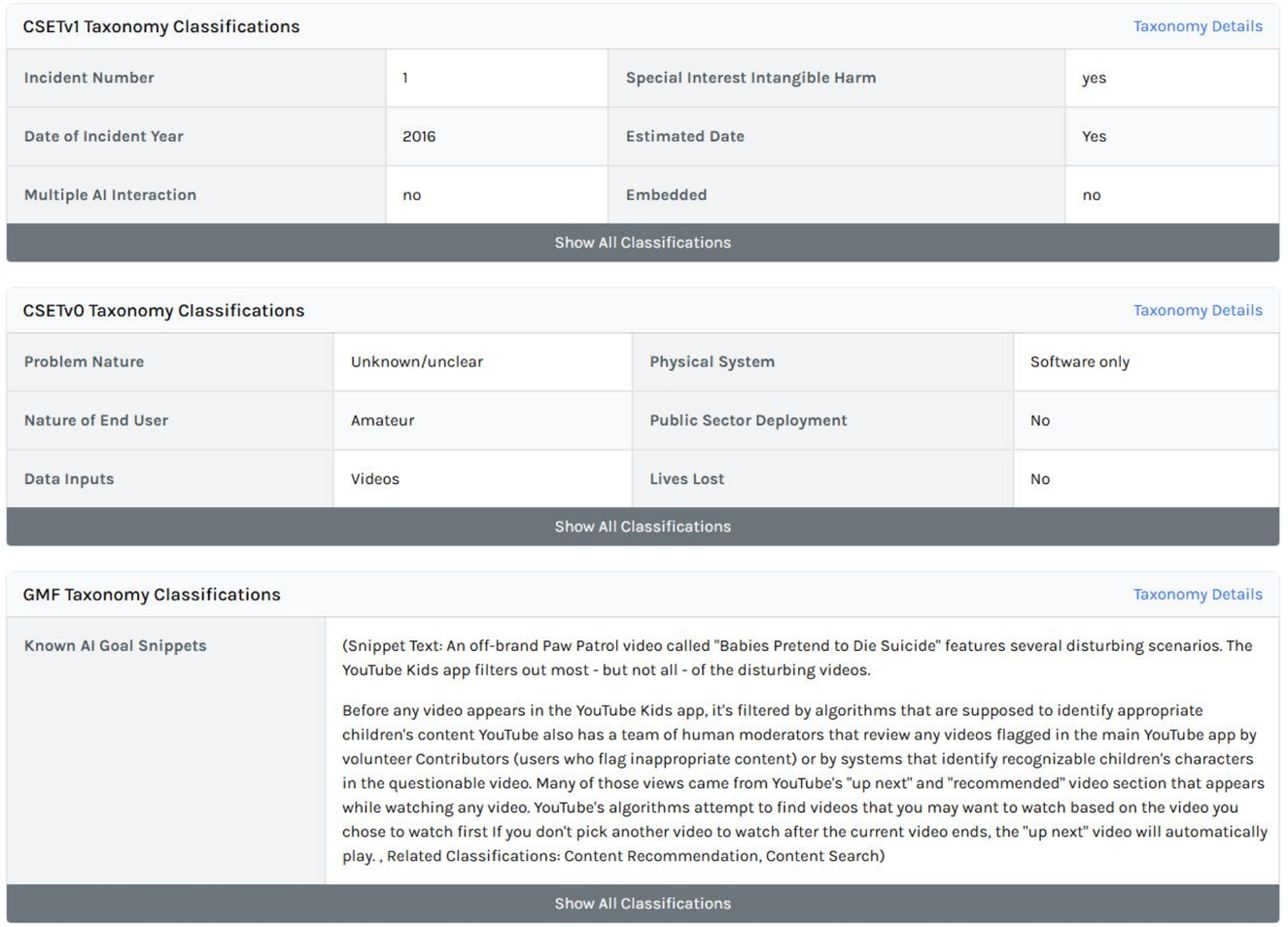
The AI Incidents Database

AI Incidents database containing 1000+ incidents.



The AI Incidents Database

Incidents usually have a short description an different taxonomies and classifications, depending on the information available.



CSETv1 Taxonomy Classifications — Taxonomy Details

Incident Number	1	Special Interest Intangible Harm	yes
Date of Incident Year	2016	Estimated Date	Yes
Multiple AI Interaction	no	Embedded	no

Show All Classifications

CSETv0 Taxonomy Classifications — Taxonomy Details

Problem Nature	Unknown/unclear	Physical System	Software only
Nature of End User	Amateur	Public Sector Deployment	No
Data Inputs	Videos	Lives Lost	No

Show All Classifications

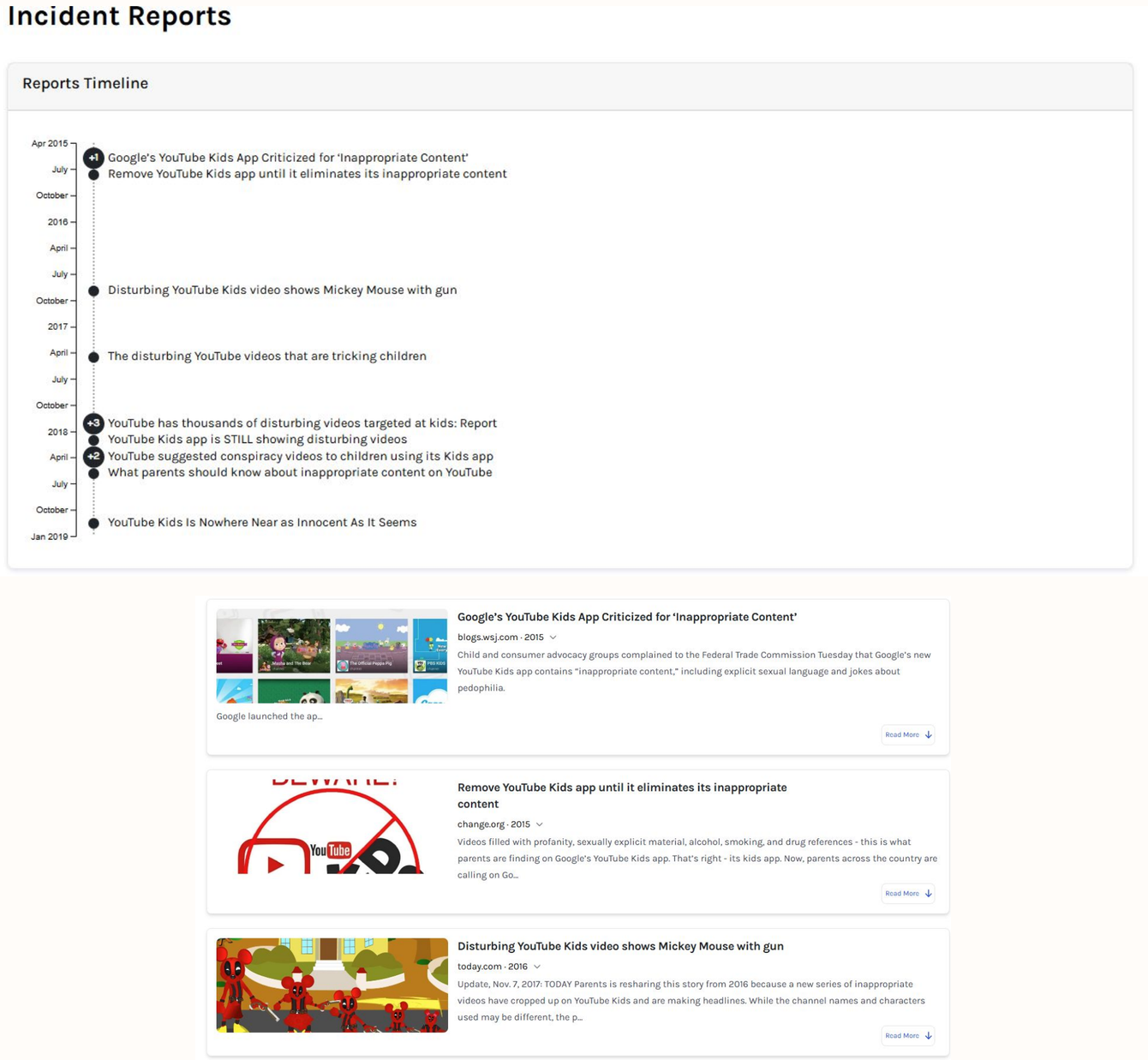
GMF Taxonomy Classifications — Taxonomy Details

Known AI Goal Snippets	(Snippet Text: An off-brand Paw Patrol video called "Babies Pretend to Die Suicide" features several disturbing scenarios. The YouTube Kids app filters out most - but not all - of the disturbing videos. Before any video appears in the YouTube Kids app, it's filtered by algorithms that are supposed to identify appropriate children's content YouTube also has a team of human moderators that review any videos flagged in the main YouTube app by volunteer Contributors (users who flag inappropriate content) or by systems that identify recognizable children's characters in the questionable video. Many of those views came from YouTube's "up next" and "recommended" video section that appears while watching any video. YouTube's algorithms attempt to find videos that you may want to watch based on the video you chose to watch first If you don't pick another video to watch after the current video ends, the "up next" video will automatically play. , Related Classifications: Content Recommendation, Content Search)

Show All Classifications

The AI Incidents Database

For each incident, there are a number of reports in a timeline an related news.



How can existing LLM interfaces support policymakers in relating AI incident to existing regulations to inform policy making?

Prompt

Based only on the information contained in the list of URLs, each of which corresponds to an AI incident, and the text of the EU AI Act, which is uploaded, perform the following analysis for each AI incident described:

1. Identify the Incident: Briefly state which incident you are analyzing (e.g., "Incident involving autonomous vehicle described in Source Document A").
2. Relevant EU AI Act Articles: List the specific articles from the provided EU AI Act text that appear most relevant to this incident.
3. Explanation of Relevance: For each listed article, explain precisely why it is relevant. Connect specific details of the incident (as found in the sources) to the specific language or requirements of the article (as found in the AI Act source). Quote or closely paraphrase key phrases from both the incident description and the Act text to support your reasoning.
4. Potential Implications & Breaches under the Act: Based on the incident description and the relevant articles, discuss potential implications according to the text of the EU AI Act. This could include: Potential risk classification (e.g., appears to meet criteria for high-risk, prohibited AI). Specific obligations that might apply or appear breached according to the text (e.g., data governance, transparency, human oversight requirements mentioned in Article X). Potential alignment or conflict with the Act's stated requirements or prohibitions.
5. Basis and Confidence in Textual Links: Briefly state how strongly the source texts support this connection. Mention if the analysis relies on direct textual links, strong thematic overlap, or interpretation based on the provided text. Note any significant ambiguities or missing information in the sources that affect this analysis. Do not provide definitive legal judgments or predict specific enforcement actions.

Please structure the output clearly, using headings for each incident analyzed.

AI Incidents URLs:
https://incidentdatabase.ai/cite/###/

Four models [Chat Interface]

- Gemini 2.5 (Google)
- NotebookLM (Google)
- ChatGPT 4.5 (OpenAI)
- Claude 3.7 (Anthropic)

Three types of run

- 1 incident
- 5 incidents
- 10 incidents

Up to 5 tries

When the prompt failed to provide an answer, we prompted again up until 5 times.

Results

Incident 658

Gemini (1)

Article 6(2) and Annex III, point 5(b): Classification of high-risk AI systems, specifically systems intended for evaluating creditworthiness or credit scores of natural persons.
Article 9: Requirements for a risk management system throughout the lifecycle of a high-risk AI system.
Article 10: Requirements concerning data and data governance, including the quality of training, validation, and testing data sets and measures to mitigate bias.
Article 14: Requirements for human oversight of high-risk AI systems.
Article 15: Requirements for accuracy, robustness, and cybersecurity.
Article 26: Obligations of deployers of high-risk AI systems, including using the system according to instructions and monitoring its operation.
Article 27: Requirement for certain deployers (potentially including financial institutions providing essential services like credit) to conduct a fundamental rights impact assessment before deploying a high-risk system.
Recital 70: Discusses the need to process special category data under specific conditions for bias detection and correction in high-risk AI systems.

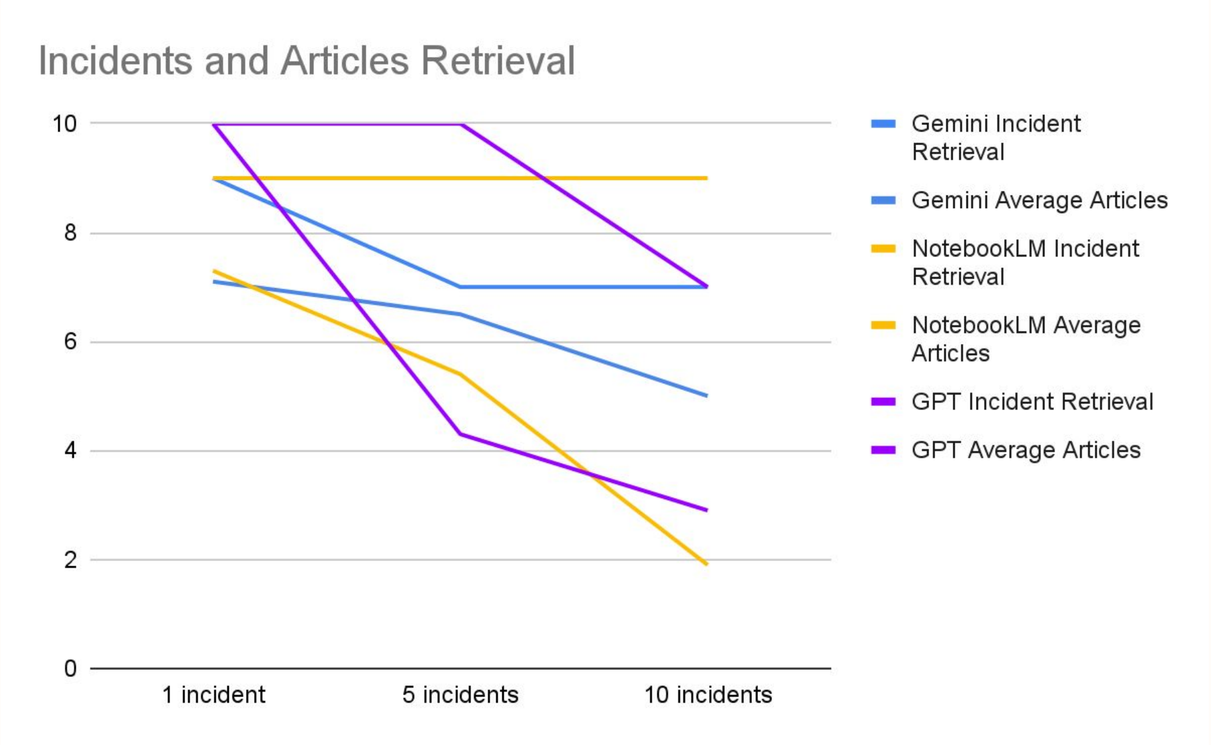
NotebookLM (1)

Article 3(1)
Article 6(2) and Annex III, point 3
Article 9(2)
Article 10(1)
Article 26(4)
Article 54(c)(ii)

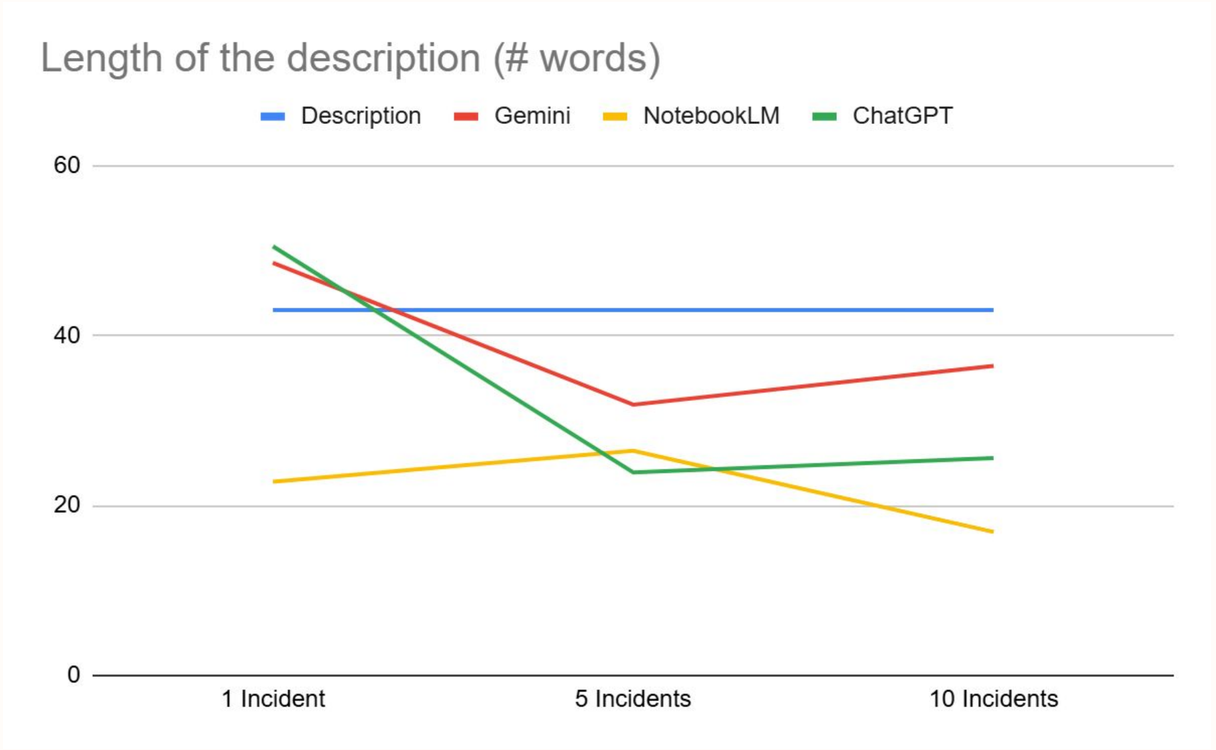
ChatGPT (1)

Article 6: Classification of high-risk AI systems
Annex III, Section 5: High-risk AI systems in access to essential private services and public services and benefits
Article 9: Risk management system
Article 10: Data and data governance
Article 13: Transparency and provision of information to users
Article 14: Human oversight
Article 15: Accuracy, robustness, and cybersecurity
Article 16: Obligations of providers of high-risk AI systems
Article 17: Quality management system
Article 19: Post-market monitoring
Article 20: Reporting of serious incidents and malfunctioning

Multiple incidents analysis yielded less incidents retrieved and an average lower number of articles identified



Multiple incidents analysis yielded less detailed descriptions in ChatGPT and Gemini, while NotebookLM always provided short descriptions



Incident 341

Nissan's Automatic Emergency Braking (AEB) feature was reported in a series of complaints for false positives and abrupt braking behaviors, endangering car occupants and traffic participants.

Gemini

- **1 Incident:** The incident being analyzed is Incident 341 from the AI Incident Database, concerning Nissan's Automatic Emergency Braking (AEB) system experiencing false positives, leading to abrupt and unnecessary braking, which posed risks to drivers and other traffic participants.
- **5 incidents:** This incident involves Nissan's Automatic Emergency Braking (AEB) system, which reportedly generated false positives and engaged in abrupt braking, posing risks to drivers and traffic participants.
- **10 incidents:** This incident involves Nissan's Automatic Emergency Braking (AEB) system in certain vehicle models, which reportedly generated false positives, leading to sudden and unnecessary braking, endangering occupants and other road users. Minor collisions and injuries were allegedly reported.

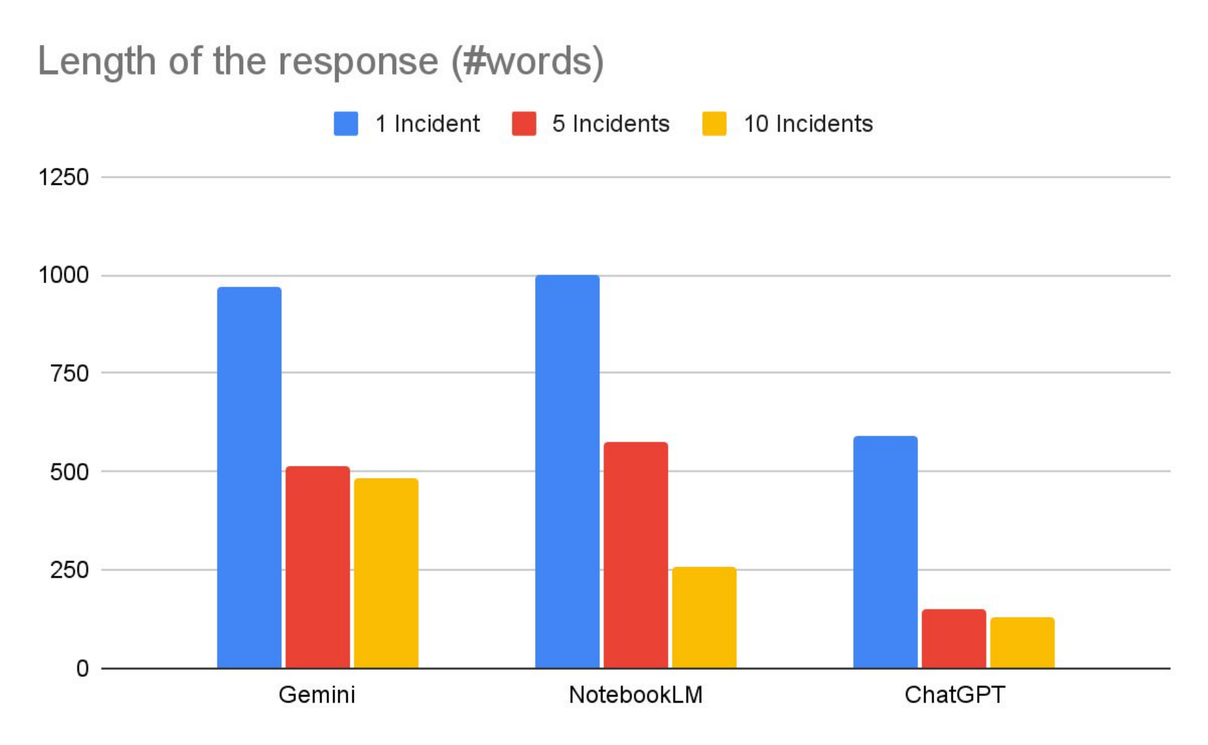
NotebookLM

- **1 Incident:** This analysis concerns "Incident 341: Nissan's 'Automatic Emergency Braking' False Positives Posed Traffic Risks to Drivers," which describes issues with Nissan's Automatic Emergency Braking (AEB) system resulting in false activations and potential collisions.
- **5 incidents:** This incident, described in "Incident 341: Nissan's 'Automatic Emergency Braking' False Positives Posed Traffic Risks to Drivers", involves Nissan's Automatic Emergency Braking (AEB) system allegedly triggering false positives, posing traffic risks to drivers and reportedly leading to minor collisions and injuries.
- **10 incidents:** Incident 341 describes how false positives in Nissan's "Automatic Emergency Braking" (AEB) system posed traffic risks to drivers.

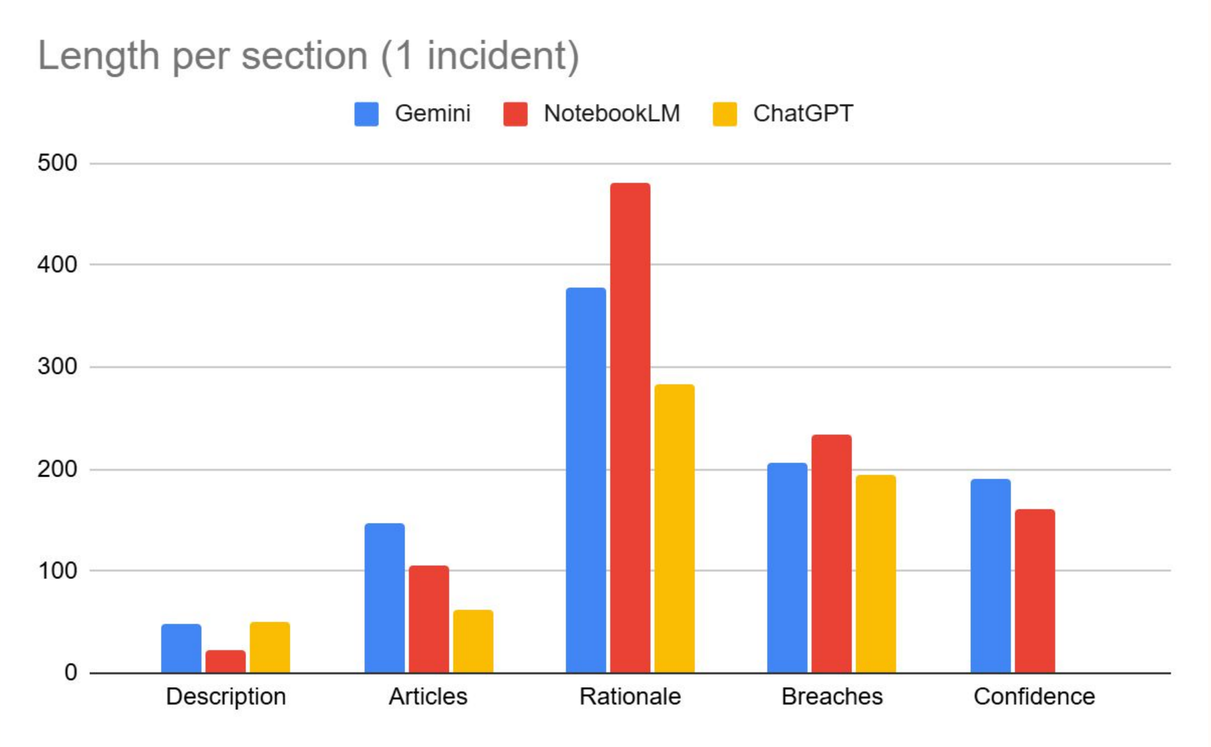
ChatGPT

- **1 Incident:** Nissan's "Automatic Emergency Braking" (AEB) system was reported to cause false positives, leading to abrupt and unnecessary braking. These incidents resulted in traffic risks, minor collisions, and injuries. Nissan acknowledged at least 30 minor collisions and 10 minor injuries related to the AEB system.
- **5 incidents:** Nissan's Automatic Emergency Braking (AEB) system reportedly experienced false positives, leading to abrupt braking behaviors that endangered car occupants and other traffic participants.
- **10 incidents:** Nissan's AEB system was reported to cause false positives and abrupt braking behaviors, leading to minor collisions and injuries.

Multiple incidents analysis decreased the overall details on the response for each incident



ChatGPT was less detailed than Gemini and NotebookLM across sections



Incident 726

Gemini (1)

Article 6 (Classification): This article is relevant because the Cruise autonomous vehicle system likely qualifies as a high-risk AI system. Autonomous vehicles are covered by Union harmonisation legislation listed in Annex I (e.g., Regulation (EU) 2018/858). If the AI system is a safety component of such a product, or the product itself, requiring third-party conformity assessment under that legislation, it's classified as high-risk under Article 6(1). Additionally, AI systems intended as safety components in the management of road traffic are listed as high-risk in Annex III, point 2.

Article 9 (Risk Management): This requires providers to implement a continuous risk management process covering the system's lifecycle, including identifying risks during intended use and reasonably foreseeable misuse. The incident, specifically the vehicle's action of attempting to pull over while dragging the pedestrian, suggests a potential failure in identifying, evaluating, or mitigating the risks associated with this specific, albeit unusual, scenario.

Article 14 (Human Oversight): This article mandates that high-risk AI systems be designed for effective human oversight to prevent or minimize risks. The incident raises questions about the design and effectiveness of any human oversight mechanisms (remote or otherwise) associated with the Cruise vehicle at the time, and whether they could have prevented the dragging action.

Article 15 (Robustness): High-risk systems must be resilient to errors, faults, or inconsistencies, especially those arising from interaction with the environment. The system's failure to respond safely after the initial impact (attempting to pull over instead of remaining stationary or performing a safer action) points to a potential lack of robustness in handling unexpected or edge-case scenarios.

Article 16 (Provider Obligations): Cruise, as the provider, has obligations including ensuring compliance with requirements (like risk management and robustness), taking corrective actions, and cooperating with authorities. The suspension of its permits and subsequent settlement relate to these obligations.

Article 20 (Corrective Actions): If a provider considers or has reason to consider its system is not in conformity, it must take immediate corrective actions (withdrawal, recall, etc.) and inform relevant parties. Cruise's actions following the incident, including halting operations, appear related to this requirement.

Article 72 (Post-Market Monitoring): Providers must collect and analyze data on the performance of their systems throughout their lifetime to evaluate ongoing compliance. This incident would be critical data for Cruise's post-market monitoring system.

Article 73 (Serious Incident Reporting): Providers must report serious incidents (defined in Article 3(49) to include death or serious harm to health) to the relevant market surveillance authorities without undue delay. The incident, resulting in life-threatening injuries, clearly falls under this reporting requirement.

Incident 726

NotebookLM (1)

Article 9 and Annex III, Point 7: The Cruise robot taxi, being an ""AI system intended to be used as a safety component in the management and operation of road traffic and the movement of persons,"" as indicated by its function as a self-driving vehicle moving people, likely falls under the definition of a high-risk AI system according to Annex III, Point 7 of the EU AI Act. This classification is triggered by the system's intended purpose and the potential for significant harm to individuals. The incident where a ""pedestrian was... dragged 20 feet by the Cruise vehicle""1 underscores the potential for harm associated with its operation in road traffic.

Article 10: Given the likely classification as high-risk, Article 10 becomes relevant as it mandates that providers and users of such systems must comply with the obligations set out in Chapter III, Section 2. Cruise, as the developer and deployer (""Alleged: Cruise developed and deployed an AI system""2), would be considered a provider.

Article 11 (Risk management system): The incident suggests a potential failure or limitation in the risk management system of the Cruise AI. A robust system should ideally anticipate and mitigate risks leading to pedestrian accidents. The fact that the incident occurred and led to ""increased regulatory scrutiny""1 might indicate shortcomings in the implemented risk management processes.

Article 13 (Data and data governance): The performance of autonomous vehicles is heavily reliant on the data used for training and operation. The incident could raise questions about the quality, representativeness, and governance of the data used by Cruise. For instance, was the system adequately trained on scenarios involving pedestrians, including those already involved in accidents?

Article 14 (Technical documentation): The EU AI Act requires comprehensive technical documentation for high-risk AI systems. This documentation should detail the system's functionalities, safety mechanisms, and limitations. The incident might prompt scrutiny of Cruise's technical documentation to assess whether it adequately addressed the risks associated with pedestrian interactions.

Article 15 (Record-keeping): The Act mandates the keeping of logs to ensure traceability of the high-risk AI system's operation. These records could be crucial in understanding the sequence of events leading to the pedestrian being dragged and in assessing the system's behavior.

Article 16 (Transparency and provision of information to users): While the ""user"" in this context might be debated (passenger or the general public interacting with the vehicle), the principle of transparency is fundamental. Information about the system's capabilities, limitations, and how it is designed to interact safely with vulnerable road users might be relevant.

Article 18 (Human oversight): Although the Cruise vehicle is autonomous, the EU AI Act emphasizes the need for appropriate human oversight mechanisms for high-risk AI systems. The incident could raise questions about the adequacy of remote monitoring or intervention capabilities and protocols.

Article 19 (Accuracy, robustness and cybersecurity): The fact that the vehicle dragged a pedestrian after an initial collision suggests a potential lack of robustness in its decision-making process in a complex real-world scenario. The system's ability to accurately perceive the situation and react appropriately seems to have been compromised.

Article 63 (if applicable): Depending on the nature of the AI model used by Cruise, if it qualifies as a ""general-purpose AI model"" under the Act, Cruise might also have obligations related to providing documentation about the model and its foreseeable limitations.

Incident 726

ChatGPT

Article 6: This article classifies AI systems that are safety components of products as high-risk. The Cruise robotaxi's autonomous driving system functions as a safety component, directly influencing the vehicle's operation and passenger safety.
Annex III, Section 1: Specifies that AI systems intended to be used as safety components in the management and operation of road traffic are high-risk. The robotaxi's AI system, responsible for navigation and obstacle detection, falls under this category.
Article 9: Mandates a risk management system throughout the AI system's lifecycle. The failure to detect and respond appropriately to the pedestrian indicates potential shortcomings in identifying and mitigating risks associated with the autonomous driving system.
Article 10: Requires data governance practices to ensure the quality and relevance of datasets. Inadequate or biased training data could have contributed to the system's inability to recognize the pedestrian, highlighting issues in data management.
Article 15: Stipulates that high-risk AI systems must be accurate, robust, and secure. The incident suggests deficiencies in the system's accuracy and robustness, particularly in complex, real-world scenarios involving unexpected obstacles.
Article 16: Outlines the obligations of providers, including ensuring compliance with the AI Act and maintaining technical documentation. Cruise, as the provider, is responsible for demonstrating that the robotaxi's AI system meets these requirements.
Article 17: Requires a quality management system to ensure compliance with the AI Act. The incident raises questions about the effectiveness of Cruise's quality management processes in identifying and addressing system vulnerabilities.
Article 19: Obliges providers to establish a post-market monitoring system to collect, document, and analyze data on the performance of AI systems. The occurrence of such a severe incident suggests potential gaps in ongoing monitoring and performance evaluation.
Article 20: Requires reporting of serious incidents and malfunctioning to the relevant authorities. Given the severity of the incident, timely reporting and transparency are critical, and any delays or omissions could constitute a breach of this requirement.

ChatGPT did not provided the explanation on its confidence in the analysis

Ej. Incident 204

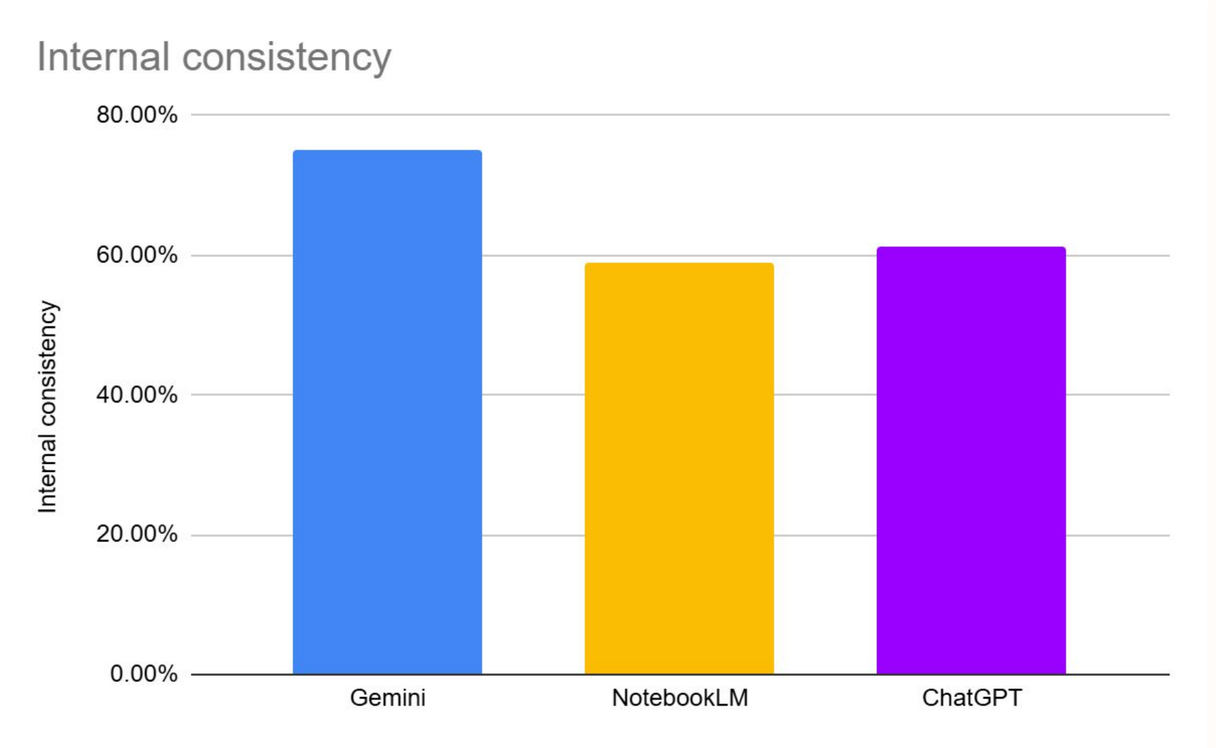
Gemini

The link between the incident and the classification as high-risk under Annex III, point 4(b) relies on a direct textual link and strong thematic overlap, as the incident involves AI used for evaluating employee behaviour potentially leading to termination. Confidence here is high.
The connection to prohibited practices under Article 5(1)(b) and 5(1)(c) relies more on interpretation based on the provided text and strong thematic overlap. The incident description doesn't explicitly use terms like ""vulnerability exploitation"" or ""social scoring,"" requiring interpretation of whether the system's function fits these legal definitions. Confidence here is medium.
Connections to breaches of specific high-risk requirements (Articles 10, 13, 14, 15) [Pages: 57-61] and deployer obligations (Article 26(7)) [Page: 68] are based on strong thematic overlap and interpretation. The incident details align closely with the risks these articles aim to mitigate (bias, lack of transparency, lack of oversight, inaccuracy, failure to inform). Confidence here is medium to high.
Ambiguities/Missing Information: The analysis is limited by the information provided in the incident summary. Key missing details include the precise algorithm workings, the exact data used (beyond ""online footprints""), the specific criteria defining ""resignation risk,"" the level of human involvement in the final firing decision, and whether the employee belonged to a legally defined vulnerable group. Zhihu's denial of using such a system also introduces ambiguity. This analysis assumes the alleged facts as presented in the incident report for the purpose of applying the Act's text. No definitive legal judgment is made.

NotebookLM

The link to Article 6(2) and Annex III, point 1 is strong as the description of the AI system's alleged use directly matches the categories of high-risk AI systems listed.
The relevance of Articles 9, 10, 12, 26, 27, and 54 is based on strong thematic overlap with the incident described and the obligations and requirements outlined in these articles for high-risk AI systems. If the system is indeed high-risk, these articles would be directly applicable.
Ambiguities and Missing Information: The provided incident description is quite brief and lacks crucial details. For instance, it does not specify if Zhihu is based in the EU, which would determine the direct applicability of the EU AI Act12. It also doesn't provide information about the AI system's functionality, the data it uses, the level of human involvement in the decision-making process, or whether a fundamental rights impact assessment was conducted. These ambiguities make it difficult to definitively determine breaches of specific articles. The analysis relies on interpreting the incident based on the limited information available and connecting it to the general principles and requirements outlined in the EU AI Act for potentially high-risk AI systems used in employment.

Gemini is the most internally consistent, but across models, consistency was only 34.89%



The models identified a baseline of general articles and a different set more related to the incident.

A	B	C	D	E	F	G	H	I	J	K
Articles	Gemini (1)	Gemini (5)	Gemini (10)	NotebookLM (1)	NotebookLM (5)	NotebookLM (10	ChatGPT (1)	ChatGPT (5)	ChatGPT (10)	**Average**
Article 3	22.22%	0.00%	14.29%	44.44%	0.00%	0.00%	0.00%	0.00%	0.00%	8.99%
Article 5	44.44%	57.14%	42.86%	33.33%	44.44%	11.11%	30.00%	30.00%	57.14%	38.94%
Article 6	88.89%	85.71%	100.00%	66.67%	55.56%	44.44%	100.00%	100.00%	57.14%	77.60%
Article 8	0.00%	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 9	55.56%	85.71%	57.14%	66.67%	44.44%	0.00%	100.00%	40.00%	42.86%	54.71%
Article 10	66.67%	85.71%	42.86%	66.67%	55.56%	33.33%	100.00%	70.00%	85.71%	67.39%
Article 11	0.00%	0.00%	0.00%	11.11%	11.11%	0.00%	0.00%	0.00%	0.00%	2.47%
Article 12	0.00%	14.29%	0.00%	11.11%	33.33%	0.00%	0.00%	0.00%	0.00%	6.53%
Article 13	33.33%	28.57%	14.29%	33.33%	11.11%	0.00%	70.00%	60.00%	57.14%	34.20%
Article 14	66.67%	85.71%	57.14%	11.11%	22.22%	0.00%	70.00%	60.00%	28.57%	44.60%
Article 15	77.78%	71.43%	57.14%	33.33%	11.11%	33.33%	90.00%	40.00%	42.86%	50.78%
Article 16	33.33%	57.14%	57.14%	11.11%	0.00%	0.00%	90.00%	0.00%	0.00%	27.64%
Article 17	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	70.00%	0.00%	42.86%	12.54%
Article 18	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	20.00%	0.00%	0.00%	3.46%
Article 19	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	90.00%	10.00%	0.00%	12.35%
Article 20	22.22%	0.00%	0.00%	11.11%	0.00%	0.00%	80.00%	10.00%	0.00%	13.70%
Article 22	0.00%	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 26	44.44%	42.86%	42.86%	33.33%	22.22%	0.00%	0.00%	0.00%	0.00%	20.63%
Article 27	22.22%	42.86%	28.57%	11.11%	11.11%	0.00%	0.00%	0.00%	0.00%	12.87%
Article 36	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 38	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 43	0.00%	0.00%	0.00%	0.00%	22.22%	0.00%	0.00%	0.00%	0.00%	2.47%
Article 50	44.44%	28.57%	28.57%	55.56%	33.33%	44.44%	0.00%	0.00%	0.00%	26.10%
Article 51	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 52	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.00%	10.00%	0.00%	2.22%
Article 53	0.00%	14.29%	0.00%	33.33%	11.11%	11.11%	0.00%	0.00%	0.00%	7.76%
Article 54	0.00%	0.00%	0.00%	22.22%	22.22%	0.00%	0.00%	0.00%	0.00%	4.94%
Article 55	0.00%	14.29%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	2.82%
Article 63	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 72	33.33%	42.86%	0.00%	22.22%	22.22%	0.00%	0.00%	0.00%	0.00%	13.40%
Article 73	33.33%	57.14%	42.86%	11.11%	22.22%	0.00%	0.00%	0.00%	0.00%	18.52%
Article 74	0.00%	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 78	0.00%	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 79	0.00%	0.00%	0.00%	22.22%	22.22%	0.00%	0.00%	0.00%	0.00%	4.94%
Article 80	0.00%	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 83	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 86	11.11%	14.29%	28.57%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	7.23%
Article 90	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 100	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	1.23%
Article 113	0.00%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	1.23%
Annex I	22.22%	42.86%	42.86%	11.11%	22.22%	11.11%	0.00%	0.00%	0.00%	16.93%
Annex III	66.67%	57.14%	57.14%	55.56%	44.44%	22.22%	80.00%	0.00%	0.00%	42.57%

Discussion

- There is a tension between ease of use (chat interface) and the sophistication of the analysis, which requires tailored software or coding.
- It is hard to uncover the factors influencing the models' capacity of fetching the correct information and analyzing it [Black Box]:
  - Privacy and cybersecurity policies.
  - Technical specifications.
  - Prompting nuances.
- Guidelines struggle with the fast-moving pace of the models (like our own research!)
- Lack of widespread standardized practices in incident detection/reporting and taxonomies to classify them pose a challenge for taking more advantage of them both through manual and automated analysis.

Limitations

Usability Tradeoff

Coding + API v/s chat interface
URLs v/s text files v/s spreadsheets
Explanations and models' (lack) of transparency

Database

Lack of coherent and stable format.
Short descriptions and heterogeneous amount of details.

Next Steps

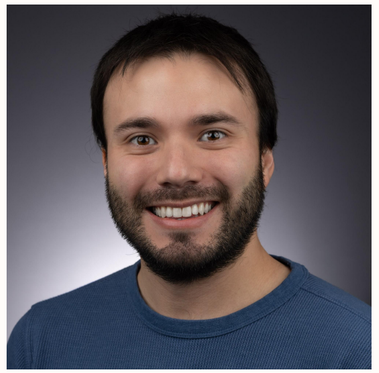
Quantitative analysis

Analyze the full AIID (1050 incidents) using APIs.
Analyze other AI incidents databases (e.g., OECD)
Test other AI regulations (e.g., Korea AI Law)

User study

Recruit lawyers / policymakers to evaluate **usability** and **quality** of the outputs.

Contact us!



Jose A. Guridi

jg2222@cornell.edu
https://joseguridi.owlstown.net/
https://www.linkedin.com/in/jaguridi/
@JguridiB on X

Eunji Emily Kim

emily.kim@gatech.edu